



Joint Anomaly Detection and Spectral Unmixing for Planetary Hyperspectral Images

Sina Nakhostin, Harold Clenet, Thomas Corpetti, Nicolas Courty

► To cite this version:

Sina Nakhostin, Harold Clenet, Thomas Corpetti, Nicolas Courty. Joint Anomaly Detection and Spectral Unmixing for Planetary Hyperspectral Images. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54 (12), pp.6879 - 6894. 10.1109/TGRS.2016.2586188 . hal-01377230

HAL Id: hal-01377230

<https://hal.science/hal-01377230>

Submitted on 6 Oct 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Joint Anomaly Detection and Spectral Unmixing for Planetary Hyperspectral Images

Sina Nakhostin, Harold Clenet, Thomas Corpetti, and Nicolas Courty

Abstract—Hyperspectral images are commonly used in the context of planetary exploration, especially for the analysis of the composition of planets. As several instruments have been sent throughout the Solar System, a huge quantity of data is getting available for the research community. Among classical problems in the analysis of hyperspectral images, a crucial one is unsupervised non-linear spectral unmixing, which aims at estimating the spectral signatures of elementary materials and determining their relative contribution at a sub-pixel level. While the unmixing problem is well studied for earth observation, some of the traditional problems encountered with earth images are somehow magnified in planetary exploration. Among them, large image sizes, strong non-linearities in the mixing (often different from those found in the earth images) and presence of anomalies are usually impairing the unmixing algorithms. This paper presents a new method that scales favorably with the problem posed by this analysis. It performs an unsupervised unmixing jointly with anomaly detection capacities, and has a global linear complexity. Non linearities are handled by decomposing the hyperspectral data on an overcomplete set of spectra, combined with a specific sparse projection, which guarantees the interpretability of the analysis. A theoretical study is proposed on synthetic datasets, and results are presented over the challenging 4-Vesta asteroid dataset.

Index Terms—Planetary Hyperspectral Unmixing, Anomaly Detection, Manifold Learning, Non-negative Matrix Factorization, Overcomplete Dictionary, Kernel based Learning.

I. INTRODUCTION

HYPERSPECTRAL (HS) remotely sensed images are of prime interest in many scientific fields, since they enable us to assess a dense spectrum (generally composed of several hundreds of contiguous electromagnetic wavelengths) in each pixel, making possible the identification of various materials composing the scene [1]. For this reason, hyperspectral data are vastly being used in earth observation (identifying land cover, crops, ... see for example [2], [3]) as well as planetary exploration (identification of minerals and rocks [4]). This latter application is the one we focus in this paper.

To understand the processes that drive the formation and evolution of planets, it is crucial to characterize their surface's composition. Visible-near infrared reflectance spectroscopy has long been recognized as an extremely powerful tool to

achieve this objective as it allows to determine rocks composition at both local and global scales [5]. Imaging spectroscopy therefore, has been rapidly grown in planetary exploration, and very soon dedicated instruments were sent toward other planets in our Solar System [6]. Number of pioneering instruments have followed resulting in huge amount of data available for the scientific community.

From the computational point of view however, the processing of hyperspectral images is particularly challenging when considering planetary exploration, since unlike the earth, no ground truth nor field campaigns exist except few particular cases (i.e. Apollo landing sites for the Moon and rovers/landers locations for Mars). Moreover, despite continuous spectral and spatial resolution improvements of HS images, they still suffer from a considerable low spatial resolution to observe every single material. As a consequence a pixel generally contains mixed information of several elements which are of key interest when considering planetary processes, especially magmatic rocks (i.e. rocks made of the minerals *olivine* and *pyroxenes*) that are used to evidence igneous processes typical of the crust or mantle. It should also be outlined that in addition to observation conditions (e.g. illumination, shadows) and to the nature of the surface itself (e.g. roughness, composition) which create non-linearities (as *multiple scattering* [7] or *intimate mixing* [8]), olivine and pyroxenes can have a wide range of chemical compositions, leading for each of them to a non-linear spectral response (e.g. [9] and references therein). Combined to this effect, rocks are themselves a combination of minerals, implying to decompose the different spectra not on a basis of pure materials but rather as a combination of mixtures of materials.

Apart from these problems related to the specificity of hyperspectral data and planetary contexts, let us also note that *anomalies* often occur in the acquisition process of hyperspectral images. Anomaly is referred to any pixel whose signature is considerably different from its surrounding background. Several factors give rise to anomalies : 1) Natural degradation due to time spent in space (especially regarding missions far away from the Solar System) where unrealistic responses can appear progressively and randomly ; 2) Varying natures of sensors used to cover the whole wavelength range which generate misregistrations ; 3) Correction of atmospheric effects where for some planetary objects (e.g. Mars, Titan) residues can affect spectra in specific wavelengths.

Although efficient pre-launch calibration procedures exist, they are not sufficient to perfectly clean data. All these reasons give rise to the necessity of efficient non-linear *unmixing* procedures, robust to outliers, able to decompose mixed pixels

S. Nakhostin and N. Courty are with the IRISA laboratory, Université de Bretagne-Sud, Vannes, 56000 France e-mail: (sina.nakhostin@irisa.fr; nicolas.courty@irisa.fr).

T. Corpetti is with the CNRS (IRISA, LETG RENNES COS-TEL) laboratory, Université Rennes 2, Rennes, 35043 France e-mail: (thomas.corpetti@irisa.fr).

H. Clenet is with Laboratoire de sciences de la Terre et des planètes, EPFL SB-ICMP-EPSL, Lausanne, CH-1015 Switzerland e-mail: (harold.clenet@epfl.ch).

Manuscript received June 28, 2016;

into a number of *pure* reflectance spectra called *endmembers*. This problem refers to an inverse procedure to determine endmembers and estimate *fractional abundances* for each pixel [10].

Traditionally, anomaly detection and spectral unmixing have taken as separate subjects. A number of families of unmixing techniques which aim at recovering abundances of endmembers directly from the image (or from a separate dictionary of acquired signatures) have been presented. Despite efficient and sound existing methods (see [11]–[13] for an overview), most techniques still suffer from degraded performances in presence of anomalies. In such situations, endmembers are often being incorrectly estimated and consequently, associated fractional abundances are less meaningful. Therefore, together with unmixing, an accurate detection of anomalies is crucial. This constitutes the aim of this paper. Before describing our methodological contribution, we now discuss into more details about unmixing and anomaly detection.

A. Hyperspectral Unmixing

1) Overview:

A large number of algorithms have been proposed to deal with HS unmixing (see [11], [12] for a comprehensive survey). These algorithms, aim either at estimating endmembers and the associated mixing abundances (unsupervised methods) or they focus only on estimating abundance coefficients on the assumption that endmembers are known (supervised methods). The unmixing methods can also be classified based on their a priori assumption on *Linearity* or *Non-linearity* of the mixing procedure. Though unmixing algorithms with linear assumption, show relatively good balance between computational complexity and acceptable accuracy, the underlying linear hypothesis is violated in many practical cases [11]. Non-linear techniques have thus been developed to cope with this difficulty including *intimate mixtures*, *bilinear models* and other physics-based nonlinear mixing models in order to give an approximate formulation of the real mixing process.

For intimate mixtures, a number of models based on the modeling of *radiative transfer* [14] have been derived. The main advantage of such approaches is the interpretability of the mixing process, since they involve real physical quantities such as grain size, roughness, single-scattering albedo, etc. However, such physical quantities, most of the time depend on unknown parameters, related to photometry and optical constants which yield a difficult estimation process. To relax the dependency on physical parameters, analytical formulations based on *bilinear models* have been successfully applied to represent the multiple scattering phenomenon which is the interactions of light with several materials. Efficient associated techniques can be found in [15]–[18].

It worths mentioning that although many unmixing algorithms assume one type of non-linearity in the mixing process, some of them (see [19] as an example) do not take any assumption on the type of nonlinearity in this process (as those working on a *feature space* through the *kernel-trick*, as

detailed in the following paragraph). The current work belongs to this family of methods. In the following we discuss about endmember and abundance estimation.

2) Estimating endmembers and abundances:

To estimate the pure elements, also called endmembers, a number of methods assume that at least one pure instance per endmember is included in the data [20]–[22]. Various techniques exist to extract them. They are either based on a pixel purity index [20] or on the identification of the simplex that encompasses data [23], [24]. This latter idea exploits the property that a linear combination of some points is included in the simplex, issued from these points; as a consequence extreme points of the simplex correspond to endmembers. N-Finder [21] is for example, issued from this representation since the maximum volume of a simplex is defined by the purest pixels in a dataset. Techniques based on non-negative matrix factorization (NMF) [25]–[28] exploit also this property and the simplex is sometimes computed through geodesic distances [29] or in *Reproducing Kernel Hilbert Space* (RKHS) [30]. This *kernel-trick* procedure enables to project data in a *feature space* (generally of high dimension), where the linearity assumption is more relevant. This efficient strategy has enabled to tackle a number of unsupervised, non-linear unmixing processes (see [7], [8], [31]–[33]). Finally let us point out VCA (Vertex Component Analysis [22]) which is based on the idea that affine transformation of a simplex is also a simplex. According to this method, endmembers are being estimated iteratively by projecting data onto the direction which is orthogonal to the subspace spanned by the previously found endmembers, until the preset number of endmembers is reached. Although efficient, all these techniques assume the presence of pure elements in the image which is not guaranteed in practice. Recent alternative approaches, based on *over complete dictionaries*, have relaxed this assumption [34], [35]. The general idea is to construct a set/dictionary of endmembers/atoms larger than the effective number of pure materials. This mainly enables to deal with the internal heterogeneity inside a particular material (which should corresponds in only one endmember in classical approaches). As will be shown in section II, our method relies on such geometrical overcomplete dictionary extraction.

Once endmembers known, an optimization procedure is often used to estimate the abundances in each pixel. It usually takes the form $\hat{\mathbf{g}}_i = \arg \min_{\mathbf{g}} \|\mathbf{x}_i - \varphi(\mathbf{E}, \mathbf{g})\|_2^2$ where \mathbf{E} represents the endmembers and $\varphi(\mathbf{E}, \cdot)$ is a nonlinear function that links the vector of abundances \mathbf{g} to the current pixel \mathbf{x}_i to unmix. Various mixing processes φ and penalization norms have been proposed (see for example [7], [15], [17]).

Let us now discuss about anomaly detection.

B. Anomaly detection

As mentioned above, anomalies are likely to be present in planetary hyperspectral data. Although noise can more or less be efficiently managed with some dimensionality reduction and denoising techniques, anomalies, because of their coherent structure, still affect their precision. All techniques, relying on endmember estimation directly from the image are sensitive

to anomalies. For example, strategies based on extraction of a simplex will embed anomalies that are most of the time outside the simplex [22], [36], [37]. The identification of these existing anomalies is then of prime importance.

Many anomaly detection techniques rely on statistical approaches where the goal is to identify among data, points with strongly different statistics w.r.t the majority of data [1], [38], these latter being viewed in a local context (local neighborhood of the current test pixel) or global one (the whole scene). On this basis, more or less advanced approaches have been proposed as for example the use of a Gaussian Mixture Model (GMM) [39] or Gaussian Markov Random Field [40] to model the nominal distribution of data.

Let us finally outline that other approaches try to represent data in other basis to better identify outliers, as in [41] where a Discrete Wavelet Transform is used to represent points whereas a kernel-PCA is exploited in [42] and [43].

C. Summary, goals of the paper

In this introduction we have discussed the necessity of both unmixing HS images and the detection of anomalies. Also interesting techniques have cited for solving each task. However, to the best of our knowledge, the joint estimation of anomalies and unmixing hyperspectral data has not been proposed yet ¹. In this work we introduce a technique entitled SAGA+ (Sparse And Geometry Aware +) able to jointly detect anomalies and perform spectral unmixing. This *Anomaly-Detection* and *Spectral Unmixing* procedure is unsupervised and defined in RKHS. The idea consists of increasing the volume of the *manifold hull* in feature space. SAGA+, in an iterative way, increases the volume of the simplex by finding the purest pixels within the dataset, while at the same time keeping track of the rate of change of the sum of projection errors (SOPE) in order to exclude anomalies from being taken as simplex vertexes.

The paper is organized as follows: Section II introduces the *Anomaly Detection* and *Spectral Unmixing* model mentioned above and in section III, evaluation of the model and comparison results using both synthetic and real datasets are presented. Finally, section IV concludes this work with an overall conclusion and pointers to possible future extensions.

II. SAGA+ : SIMPLEX VOLUME MAXIMIZATION AND ANOMALY DETECTION

In this section the methodological contributions are presented. The spectral unmixing relies on a matrix factorization technique denoted SAGA (Sparse And Geometry Aware) and introduced in [30]. In the present contribution we add an anomaly detection and removal approach to deal with aforementioned issues, yielding an approach denoted "SAGA+". SAGA is based on the geometrical concept of finding the simplex that embeds data. This simplex is being computed in a feature space associated to a kernel. As for anomalies, they

are tracked through their lack of representational capacity in the dataset.

We first present the endmember extraction technique in II-A and subsection II-B is devoted to anomaly detection. Once endmembers extracted, abundances are estimated by projecting data onto the endmembers. This procedure is detailed in II-C. The overall process, entitled SAGA+ is summarized in II-D. Before entering into details, let us introduce some general notations:

The hyperspectral image contains D bands and N pixels $\mathbf{x}_i \in \mathbb{R}^D, i = 1, \dots, N$. Data are represented in a $D \times N$ matrix composed of N vectors $\mathbf{X} = [\mathbf{x}_1^T, \dots, \mathbf{x}_N^T]$. The unmixing is formalized as finding a representation of the form : $\mathbf{X} \approx \mathbf{E}\mathbf{G}$ where columns of $\mathbf{E} \in \mathbb{R}^{D \times \ell}$ contain the ℓ endmembers and $\mathbf{G} \in \mathbb{R}^{\ell \times N}$ is the matrix, consisting of projections of each pixel onto this endmember basis. In practice the interpretability of matrix \mathbf{G} is enhanced if each column is positive and sums to 1, since in this case, each value \mathbf{G}_{ij} contains the abundance of endmember i in pixel j . Such a Non-negative Matrix Factorization (NMF) is the basic idea on which we rely. To estimate the endmember matrix \mathbf{E} , we rely on Column Subspace Selection principles: As for each column of \mathbf{E} is selected an input pixel \mathbf{x}_i . This idea suggests that at least one sample of each endmember is present in the data (since this assumption is likely to be violated, in practice we rely on overcomplete strategies, as will be explained). The following section introduces the simplex approach to estimate matrix \mathbf{E} .

A. SAGA principles: Exact Simplex Volume Maximization for endmember extraction

As mentioned earlier, we rely on a Column Subspace Selection procedure to construct the simplex that embeds the volume generated by the data [45]. To this end we start from the idea that the bigger this simplex is, the more likely it will contain the other elements, and therefore the reconstruction error of any projection onto the new simplex will be small or null. The ℓ vertices of the simplex then correspond to the endmembers.

In this study, we also claim that choosing an overcomplete endmembers dictionary will help in performing non-linear unmixing. This point is illustrated in Figure 1. In the first case (Figure 1(a)), the spectra are living in a perfectly linear manifold. In this case, two endmembers ($E1$ and $E2$) are sufficient and any spectrum can be described as a convex combination of the two endmembers. This is referred to as the linear mixing hypothesis. However, this ideal case is usually different from reality, as illustrated on Figure 1(b), where the spectra live on a non-linear manifold. If the linear assumption is nonetheless assumed, then the spectra are projected on the 1-simplex formed by the two endmembers, resulting in an error in the corresponding abundancies. By choosing an appropriate additional endmember ($E3$, blue circle in Figure 1(c)), one can build a piece-wise linear approximation of the non-linear manifold. The unknown spectra are now projected on the corresponding 2-simplex. Note here that this projection is naturally sparse, *i.e.* for all spectra only two endmembers are required to describe perfectly the projection.

¹Unfortunately, only after submitting this draft we became aware of a very recent work [44], using Bayesian framework for joint anomaly detection and spectral unmixing.

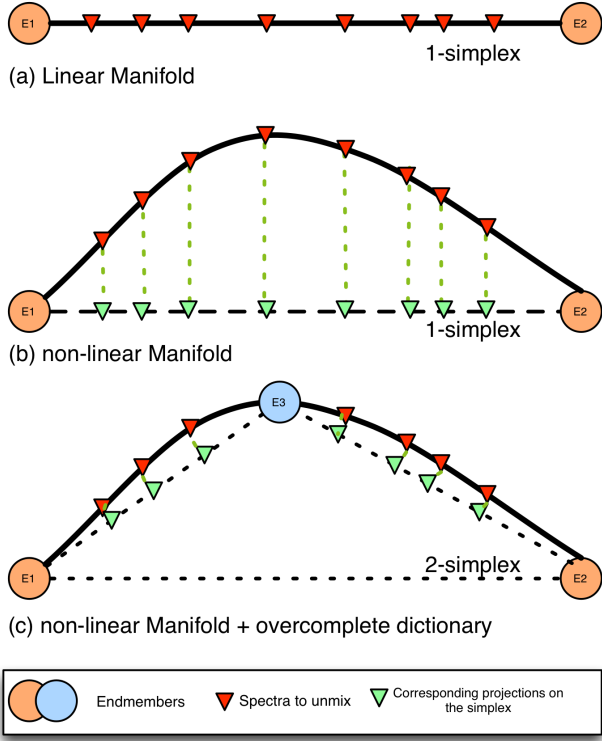


Fig. 1: Overcomplete endmember dictionary for non-linear unmixing

However this comes at the price of having one endmember that is in itself a mixture of the first two ones. There are then two cases:

- 1) Either one wants only to estimate the mixture coefficients over the two “pure” endmembers $E1$ and $E2$. Assume for the example that $E3$ is a mixture of known coefficients (0.4 and 0.6) of $E1$ and $E2$. Then for any spectrum with decomposition $\mathbf{g}_1 = 0.2$, $\mathbf{g}_2 = 0$ and $\mathbf{g}_3 = 0.8$, the corresponding abundance *w.r.t.* only $E1$ and $E2$ is now $\mathbf{g}_1 = 0.2 + 0.8 \times 0.4 = 0.52$ and $\mathbf{g}_2 = 0 + 0.8 \times 0.6 = 0.48$. More generally, in the case where λ pure endmembers are available and ℓ endmembers are chosen for the overcomplete dictionary, we suppose that we have access to a (generally sparse) linear operator \mathcal{A} of size $(\lambda \times \ell)$, with prescribed sums of rows $= 1$, that can be applied on any result of unmixing *w.r.t.* the overcomplete dictionary to get only a decomposition over pure endmembers.
- 2) However, we agree that obtaining this mixing matrix might not be a trivial task even for specialists. Nevertheless in some domains, as the one explored in this paper, specialists are used either to work with spectra obtained in laboratory that are already combination of pure chemical elements. The decomposition on the overcomplete basis, provided it can be interpreted, gives in itself interesting insights on the underlying geological process.

In addition of the overcomplete strategy, to deal with com-

plex manifolds issued from nonlinearities in \mathbf{X} , it is possible to project the data in a feature space \mathcal{H} using a projection function $\phi : \mathbb{R}^D \rightarrow \mathcal{H}$ (working in the original space is equivalent to choose the identity as projection function ϕ). We represent this projected data as: $\Phi = \phi(\mathbf{X}) = [\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \dots, \phi(\mathbf{x}_N)]$.

Under the assumption that endmembers are invariant by transformation in the feature space, the endmember extraction therefore can be written as finding an indicator matrix $\mathbf{W}^{(\ell)} \in \{0, 1\}^{N \times \ell}$ where:

$$\phi(\mathbf{X}\mathbf{W}^{(\ell)}) = \Phi\mathbf{W}^{(\ell)}. \quad (1)$$

This gives rise to a simplicial convex hull in \mathcal{H} , noted $\Delta^\ell(\phi(\mathbf{E}))$, whose pre-image is a non-simplicial hull (the manifold hull) in the original space.

The procedure first extracts the best 1-simplex Δ^1 by finding two instances from Φ that results in the minimum projection error on this basis, then iteratively increases the dimension of the simplex in \mathcal{H} . At the (p) -th iteration the algorithm selects $\phi(\mathbf{x}_i)$ such that the volume of the Δ^{p-1} -simplex is maximized, i.e.

$$i = \arg \max_q \text{Vol}(\Delta^{p-1}(\Phi\mathbf{W}^{p-1}) \cup \phi(\mathbf{x}_q)), \quad (2)$$

where the volume of Δ^p simplex in this iteration is [30]:

$$\text{Vol}(\Delta^{p-1}(\Phi\mathbf{W}^{p-1}) \cup \phi(\mathbf{x}_q)) = \frac{\text{Vol}(\Delta^{p-1}) \times \text{dist}(\phi(\mathbf{x}_q), \Phi\mathbf{W}^{p-1})}{p-1}. \quad (3)$$

In practice the optimal projection ϕ enabling to separate complex data is unknown and potentially of high dimension. Hopefully, thanks to the kernel theory [46], one can rely on kernel functions $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$ that allow to rewrite the relation $\text{dist}(\phi(\mathbf{x}_q), \Phi\mathbf{W}^{p-1})$ in a closed form:

$$\text{dist}(\phi(\mathbf{x}_q), \Phi\mathbf{W}^{(p-1)}) = 1 - (k_{x_q}^T \cdot K_{p-1}^{-1} \cdot k_{x_q}), \quad (4)$$

with K_{p-1}^{-1} is the inverse of the kernel matrix of the elements of the manifold:

$$K_{p-1}^{-1} = (\mathbf{W}^{(p-1)\top} \Phi^\top \cdot \Phi \mathbf{W}^{(p-1)})^{-1} \quad (5)$$

and k_{x_i} is a vector of length p defined as $k_{x_i} = [k(\mathbf{x}_i, \mathbf{x}_j)]_{\mathbf{x}_j \in \mathbf{X}\mathbf{W}^p}$. It is worth noting that instead of inverting the kernel matrix, in the implementation, for the sake of more numerical stability we chose to apply a Cholesky factorization on the kernel matrix at each iteration and then solve for it.

As for the choice of the kernel function, working on the original space is equivalent to choose k as the usual dot product : $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle$. Choosing a different function enables to perform the process in a feature space where the geometry of the manifold embedding \mathbf{X} can be more meaningful. Among existing kernels, the standard Gaussian

Radial Basis Function (RBF) is widely used since it enables a projection in an infinite dimensional space :

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right). \quad (6)$$

The overall procedure enables to extract endmembers by finding the optimal manifold hull composed of ℓ spectra in the feature space. This exact simplex volume maximization procedure is the cornerstone of endmember estimation in SAGA [30]. In most methods assuming a linear mixing model, the number of endmembers is exactly defined by the number of pure spectra present in the image, and cannot exceed the spectral dimension. As such, choosing a point which is already a mixture of those pure spectra would lead to a nul increase in the volume of the simplex. In the RKHS corresponding to the Gaussian Kernel, each elements of Φ are orthogonal, and therefore would lead to a positive increase of the volume. In practice, it is therefore possible to have more endmembers than the number of pure materials (and can even exceed the spectral dimension). From this point of view, the process is a dictionary learning procedure, with an overcomplete dictionary (in the sense of the original spectral space). Apart from handling implicitly the non-linearity of the mixing model, this leads to two important properties of our method:

- The spectra can be decomposed over elements that are also combination of pure elements. This can ease the interpretation of the unmixing for specialists through the obtention of more physically realistic spectra, and alleviate to some extent the purity assumption which states that at least one pixel of each material should be present in the image;
- Choosing an overcomplete endmember set can help in handling the intra-class variability of each element, as similar versions of the same material (but with different spectral signatures, as it is the case with shadows for instance), can be efficiently exploited.

In practice, this also requires some sparsity assumption in the mixing coefficients, which is guaranteed by our projection method (see Section II-C).

B. Anomaly Detection

As mentioned above, SAGA procedure selects endmembers in a way to increase the volume of the simplex hull in the feature space. According to this strategy, in each iteration, extremal points of the dataset hull are chosen. This criterion makes the algorithm vulnerable to anomalies since they are most likely to appear in the construction of our dictionary \mathbf{E} . This issue is illustrated in the two first rows of Fig. 2. In this figure, all instances of the first row are generated from a uniform Dirichlet distribution within the 2-simplex identified by blue vertexes (note that these pure points are not included in the dataset). In the second row 10 anomalies have been added to the nominal data. On the right column, the corresponding endmembers extracted by SAGA are depicted (red dots) using both datasets. From this simple experience, it is visible that endmember extraction algorithm fails to identify the correct borders of the nominal data in presence of anomalies.

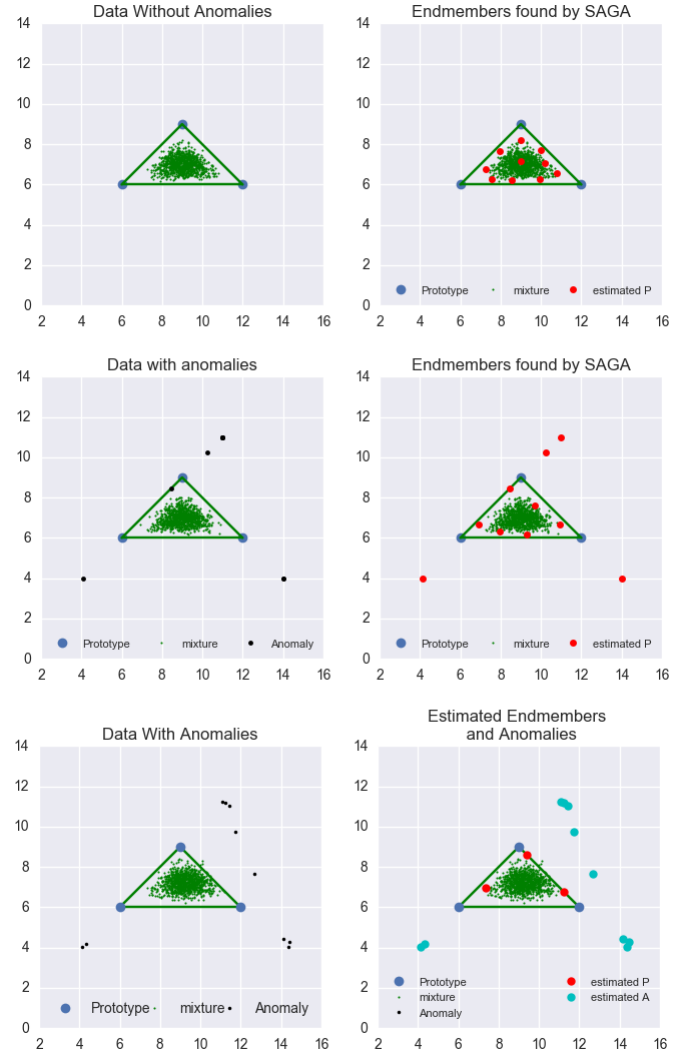


Fig. 2: **Illustration of difficulties in presence of outliers.** **First Row: endmember extraction using SAGA without anomalies:** in this situation extracted endmembers are a reliable representation of the manifold ; **Second Row: endmember extraction using SAGA in presence of anomalies:** in this situation extracted endmembers (in red) do not correspond to real borders of the data manifold ; **Third Row: endmember extraction and anomaly detection In presence of anomalies using SAGA+:** here extracted endmembers (in red) by are in accordance with borders of the manifold embedding data while the anomalies (in cyan) are properly selected.

As soon as extreme points appear, they significantly contribute in maximizing the simplex volume and are thus selected as endmembers. This behavior is observable in all unsupervised endmember extraction procedures which work based on maximization of the geometrical volume of the simplex encompassing data. To cope with this issue, the proposed anomaly detection (AD) procedure relies on the computation of the reconstruction error (in feature space) between a data sample and its perpendicular projection onto a side of the polygon of the simplex. Although similar ideas have already

been proposed (see [42] and other AD algorithms), the process here does no attempt for identifying the nominal distribution (for example based on decomposition of the eigen-structure of dataset) but we rather exploit the iterative structure of SAGA where in step p , after selection of the $p + 1$ -th endmember candidate through equation (2), we check its validity by computing the *sum-of-projection-errors* (SOPE) of all instances with respect to the new simplex Δ^{p+1} :

$$\text{SOPE}(p+1) = \frac{\left[\sum_{i=1}^N 1 - (k_{x_i}^T \cdot K_{p+1}^{-1} \cdot k_{x_i}) \right]}{N}, \quad (7)$$

where N is the number of samples and acts as the normalization constant. A perfect reconstruction gives $\text{SOPE}(p+1) = 0$, whereas its value increases when the quality of the reconstruction decreases (i.e. the simplex is not a good representation of the dataset). During the iterative procedure, when a considerable drop of SOPE is perceived after adding a new vertex to the simplex, this means that previously selected prototypes (endmember candidates of previous steps) could be anomaly and are among instances significantly different from nominal data. By this procedure one can then identify and reject them. In the third row of Fig. 2 one can see that this AD procedure, even when data are exposed to anomalies, endmember estimation remains close to a good representation of the data manifold which is in contrast to the second row of Fig.2 (without any AD process).

In order to be able to check this drop in each iteration we need to define a threshold τ . This point is not trivial and highly depends both on the distribution of nominal data and the distribution of anomalies. To illustrate this, Fig. 3 represents the evaluation of SOPE when all nominal data (green points) are represented on a unique candidate issued from a meshgrid. From this figure, one easily observes that error decreases when endmember candidates get closer to the center of the data distribution². The iso-contour corresponding to $\tau = 0.01$ in SOPE value is also represented to evaluate the associated threshold that enables to remove outlier data. A discussion related to the practical selection of τ is proposed in section III.

Once endmembers are identified (together with anomalies), the projection of each data point onto the new basis has to be performed. This process is discussed in the following.

C. Sparse Projections onto the Simplex

Let us recall that the unmixing process aims at approximating the data matrix \mathbf{X} through $\mathbf{X} \approx \mathbf{E}\mathbf{G}$, where the estimation of the endmember matrix \mathbf{E} has been presented above. As for the abundance matrix \mathbf{G} , it corresponds to the projection of \mathbf{X} onto a region bounded by \mathbf{E} . This is done, for each row of \mathbf{G} (each row being represented as $\mathbf{G}_{\bullet i}$ associated to the i th point), with a projected gradient decent :

$$\mathbf{G}_{\bullet i} = \arg \min_{\mathbf{G}_{\bullet i}} \|\phi(\mathbf{x}_i) - \Phi \mathbf{W} \mathbf{G}_{\bullet i}\|^2 \text{ s.t. } \mathbf{G}_{\bullet i} \in \Delta^\lambda. \quad (8)$$

²In order to illustrate more into details, various maps of SOPE fields along the endmember extraction process are represented in the supplemental materials.

where Δ^λ is the sub-simplex included in Δ^ℓ but composed only with λ vertexes.

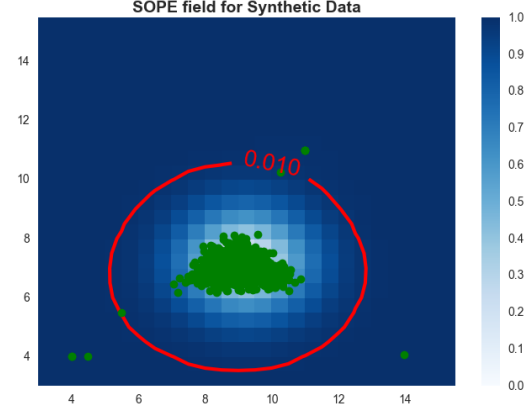


Fig. 3: **Evolution of SOPE (Sum Of Projection Errors)** when all data in green are represented through a single endmember uniformly represented on a meshgrid. One observes a decrease in SOPE when candidates get closer to the nominal distribution.

We indeed want to impose sparsity on each column of \mathbf{G} since in practice, a pixel is likely to be composed only with a subset of existing endmembers. Explicitly, imposing sparsity prevents from solutions where all endmembers have a (even small) contribution. To this end, a greedy selector and sparse projector (**GSSP**) [47] method is used not only to project the data onto the simplex but also to impose a sparsity level (through a number of non null elements given as an input parameter). In practice the coordinate values of each pixel will be sorted based on their magnitude and the λ greatest ones will be projected onto a unit simplex while the rest of the values will be set to zero.

We now have all ingredients for the complete process of anomaly detection and unmixing which is being presented.

D. Overall process

The SAGA+ procedure is presented in algorithm of Fig. 4. In a first step we compute the **manifold hull** to derive endmembers together with anomaly detection (lines 4 to 10). To initialize the process, the algorithm selects a random datum \mathbf{x}_1 (line 4), and computes the distances between \mathbf{x}_1 and all instances in \mathbf{X} (line 5). These distances are then sorted in descending order and will be feed to the Exact Simplex Volume Maximization and Anomaly Detection process to estimate the first endmember $\hat{\mathbf{e}}_1$ associated with a set of anomalies (line 6). This process is repeated ℓ times in order to extract the ℓ endmembers (lines 7 to 9) and to construct the manifold hull \mathbf{E} .

In each iteration the volume increase of the simplex is computed in RKHS w.r.t each \mathbf{x}_i . Following equations (2)–(5), this requires to inverse the kernel matrix \mathbf{K}^{-1} which is done in practice using the Cholesky decomposition. After

computation of the volume changes, the anomaly detection algorithm is called to check the trade-off between the increase of the volume of the simplex and the decrease in the sum-of-projection-errors (SOPE). This is represented in algorithm of Fig. 5.

Once the manifold hull extracted, the projection matrix \mathbf{G} is calculated using the Greedy Selector and Sparse Projector (GSSP) algorithm. This second step of SAGA+ follows the process described in [47].

The principle of Exact Simplex Volume Maximization and Anomaly Detection algorithm is described in Fig. 5. Here we check iteratively, on the basis of the SOPE criteria presented in equation (7), if successive candidates for the simplex vertexes (sorted based on their distances with the current simplex) are real endmembers or lie among outliers.

SOPE is computed by taking into account the presence of already found prototypes and in case no endmembers are already present, the sum of projection error reads $SOPE = [(\nu^\top \cdot \nu)/N] - 1$ where ν represents the vector of the similarities between the current candidate and the rest of the dataset instances.

```

1: procedure SAGA+
2: Input data
   • Data matrix  $\mathbf{X}$ 
   • Number of required endmembers  $\ell$ 
   • Threshold  $\tau$  required for SOPE (see section II-B)
   • Sparsity level  $\lambda$  for abundance matrix  $\mathbf{G}$ 
   • Kernel bandwidth  $\sigma$ 
3: Output data
   • Endmember matrix  $\mathbf{E}$ 
   • Abundance matrix  $\mathbf{G}$ 
   • Anomaly matrix  $\mathbf{A}$ 

   ▷ Step 1: Endmember Estimation (EE) and Anomaly Detection (AD):
4:   - Select a random point  $\mathbf{x}_1$ .
5:   - Compute distances w.r.t  $\mathbf{x}_1$  in RKHS (see eq. (4)–(6)).
6:   - Identify the first endmember and first outliers using Exact SiVM and AD algorithm (see description in Fig. 5)

7:   for ( $p = 2, \dots, \ell$ ) do
   - From current simplex with  $p - 1$  endmembers, compute volume changes with all instances (see eq. (3)).
   - Sort the vector of volume changes decreasingly (see eq. (2)).
   - Since first indexes are possible anomalies, SAGA+ to identify endmembers and outliers (see algorithm in Fig. 5)
8:   end for
9:   return( $\mathbf{A}, \mathbf{E}$ )

10:   ▷ Step 2: Abundance estimation via sparse projections
11:   - Apply eq.(8) and extract matrix ( $\mathbf{G}$ )
12:
13:   return ( $\mathbf{G}$ )
14: end procedure

```

Fig. 4: **Complete SAGA+ procedure:** outlier detection, end-member and associated abundance estimations.

```

1: procedure EXACT SiVM AND AD
2: Input data
   • Data matrix  $\mathbf{X}$ 
   • Current endmembers  $\hat{\mathbf{E}}$ 
   • Current outliers  $\mathbf{A}$ 
   • Vector of Sorted indexes  $\mathbf{vs}$ 
3: Output data
   • New Endmembers  $\hat{\mathbf{e}}$ 
   • New Anomalies  $\mathbf{A}$ 

4:   flag = true (Indicator of anomaly)
5:    $i = 0$ 
6:   while flag = true do (Loop on all points)
7:     if ( $\mathbf{vs}_i \notin \{\mathbf{A} \cup \hat{\mathbf{E}}\}$ ) then
8:       Compute SOPE on  $\mathbf{X}[\mathbf{vs}_i]$  Eq (7)
9:       if ( $sope < \tau$ ) then
10:        flag = false
11:       else
12:         $i++ = 1$ 
13:         $\mathbf{A} += \mathbf{x}_i$ 
14:       end if
15:     end if
16:   end while
17:   return ( $\mathbf{A}, \hat{\mathbf{E}} + \mathbf{x}_i$ ) Index of the next endmember
18: end procedure

```

Fig. 5: **Sub-procedure for endmember extraction and anomaly detection**

III. EVALUATION

In this section we evaluate on synthetic and real data, SAGA and SAGA+ procedures and compare their performances with a number of state-of-the-art unmixing procedures. In particular we consider two linear unmixing methods N-Finder [21] and VCA [22] introduced above. As mentioned, these are frequently used unsupervised methods of unmixing and their endmember extraction procedures are based on the simplex volume maximization method. This makes them suitable to be compared with the current method.

It should be mentioned however that these models assume linear mixing and do not detect outliers, therefore the comparison with our approach would be thought to some extent unfair. Nevertheless it should also be emphasized that they all follow the same logic of extending the encompassing simplex which makes them (to some degree) comparable to the logic behind SAGA. We also decided to compare our approach with G-SiVM [48] which is a non-linear procedure based on simplex volume maximization through shortest-path distances in a nearest-neighbor graph. Non-linearity is performed using kernel trick which makes this approach similar to SAGA and SAGA+ in the sense that they do not rely on any specific mixing assumption.

In order to evaluate and compare the efficiency of the procedures, we use different criteria to assess endmember extraction and anomaly detection processes. These criteria are introduced below together with experimental setups.

A. Experimental setups

1) *Evaluation criteria:* for synthetic data where ground truth is available, a set of criteria are used to measure our performances:

- To evaluate the adequacy of **estimated endmembers** $\hat{\mathbf{E}} = [\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_\ell]$ with **ground truth endmembers** $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_\ell]$, several possibilities are offered as Spectral Information Divergence (*SID*) or *ℛℬℱ* kernel distance, each of which give a measure of similarity between two signals represented as a vector. Though efficient, these criteria are sensible to the difference of endmembers in terms of magnitude while in this application, we are more interested in comparing their shape. Indeed, because of atmospheric attenuation or variation in illumination, spectral energy can be attenuated for two samples even if they belong to the same class. Thus, we rather prefer to rely on Spectral Angle Map (*SAM*) which is scale invariant. The *SAM* reads, for spectral vectors $\mathbf{x}_i = [x_{i1}, \dots, x_{iD}]^T$ and $\mathbf{x}_j = [x_{j1}, \dots, x_{jD}]^T$ with positive values:

$$\text{SAM}(\mathbf{x}_i, \mathbf{x}_j) = \cos^{-1} \left(\frac{\mathbf{x}_i \cdot \mathbf{x}_j}{\|\mathbf{x}_i\| \|\mathbf{x}_j\|} \right) \quad (9)$$

and lives in the range $[0, 1]$ (1 representing maximum dissimilarity and 0 indicating two identical signals).

In practice in order to compare two different sets of endmembers, we need a pairing procedure to associate each estimated endmember to a unique real one. This is done as finding the best permutation of estimated endmembers that gives the minimal average *SAM* values when they are compared to real ones. Therefore, two sets \mathbf{E} and $\hat{\mathbf{E}}$ have a difference $\epsilon(\mathbf{E}, \hat{\mathbf{E}})$ which reads:

$$\epsilon(\mathbf{E}, \hat{\mathbf{E}}) = \min_{\eta} \frac{1}{\ell} \sum_{i=1}^{\ell} \text{SAM}(\mathbf{e}_{\eta(i)}, \hat{\mathbf{e}}_i) \quad (10)$$

where ℓ is the number of endmembers and $\eta^\ell : \{1, \dots, \ell\} \rightarrow \{1, \dots, \ell\}$ is a function corresponding to all possible permutations.

- As for the evaluation of the anomaly detection algorithm, we rely on Cohen's Kappa Statistics [49] to measure the adequacy of the classification between outliers/endmembers. Kappa criterion, noted κ , lies between 0 (nothing is in accordance) and 1 (perfect accordance) and combines False Positive (FP), False Negative (FN), True Positive (TP) and True Negative (TN) in order to measure the inter-rater agreement of categories.
- In complement we also give an execution-time comparison among the endmember extraction procedures.

We now introduce our datasets.

2) *Synthetic data*: In order to generate synthetic datasets, we have selected 41 spectral signatures from RELAB spectral library [9]³.

From ℓ spectral signatures $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_\ell]$ ($\mathbf{e}_i \in \mathcal{R}^D$), mixed spectra have been generated based on three different mixing assumptions: Linear Mixing Model (LMM), Bilinear Mixing Model (BMM) and Highly Concentrating Model (HCM). These mixing models are well studied in the unmixing literature, though they do not fully represent the variability of mixing processes in real world hyperspectral images. The two

first models (LMM, BMM) are implemented according to the procedure in [7]. The usage of HCM model is also mentioned in some earlier works [50].

From the set of endmembers \mathbf{E} , the **Linear Mixture Model** reads :

$$\mathbf{x}_i = \mathbf{E}^T \mathbf{g}_i + \mathbf{w}_i \quad (11)$$

where \mathbf{g}_i is the abundance vector related to data \mathbf{x}_i and \mathbf{w}_i is a random noise. Each element in \mathbf{g}_i lives in the range $[0, 1]$ and their sum equals 1. The fractional abundances in LMM have been generated based on Multivariate Symmetric Dirichlet Distribution, uniformly distributed over the standard $\Delta^{\ell-1}$ simplex. More precisely for each sample \mathbf{x}_i^T , $i = 1, \dots, N$ and $\mathbf{x}_i \in \mathcal{R}^D$ the abundance vector $\mathbf{g}_i = [g_1, \dots, g_\ell]$ is generated through a Dirichlet Distribution. This latter, parametrized by vector $\alpha = [\alpha_1, \dots, \alpha_\ell]$ such that $\alpha_i > 0, \forall i \in [1, \ell]$, reads:

$$\mathbf{g}_i = \frac{\Gamma(\alpha \cdot \ell)}{\Gamma(\alpha)^\ell} \prod_{j=1}^{\ell} \mathbf{e}_j^{\alpha_j - 1} \quad (12)$$

where $\Gamma(\cdot)$ is the Gamma function. Dirichlet are generalization of the beta distribution in a multi-variate context. Such distributions are then very useful to simulate mixture models. From equation (12), using a parametrization $\alpha = [\alpha_1, \dots, \alpha_\ell]$, we generate mixtures of pure endmembers $[\mathbf{e}_1, \dots, \mathbf{e}_\ell]$ by computing abundance parameters $\mathbf{g}_i = [g_1, \dots, g_\ell]$ as the power-product between endmembers and α . This is written in the right part of equation (12). The left part is only a normalization coefficient that involves the Gamma function. In practice α is set as the vector of ones: $\alpha_i = 1, \forall i$. As the result the vector \mathbf{g}_i contains coefficients which are uniformly distributed across the ℓ vertexes of the simplex. To illustrate this, from pure endmembers visible on the top of Fig. 6(a), blue points of Fig. 6(b) represent a 2D slice (band 70 vs. band 100) of the synthetic data generated under the LMM model.

For **Bilinear Mixing Model**, the power products of reflectance is added to the linear mixing model, leading to a model of the form :

$$\mathbf{x}_i = \underbrace{\mathbf{E}^T \mathbf{g}_i}_{\text{LMM}} + \sum_{k=1}^{\ell-1} \sum_{l=k+1}^{\ell} \beta_{k,l,i} \mathbf{e}_k \odot \mathbf{e}_l + \mathbf{w}_i \quad (13)$$

where $\beta_{k,l,i}$ are new mixing coefficients and the termwise Hadamard product is defined as

$$\mathbf{e}_k \odot \mathbf{e}_l = \begin{pmatrix} e_{1,k} \\ \vdots \\ e_{D,k} \end{pmatrix} \odot \begin{pmatrix} e_{1,l} \\ \vdots \\ e_{D,l} \end{pmatrix} = \begin{pmatrix} e_{1,k} e_{1,l} \\ \vdots \\ e_{D,k} e_{D,l} \end{pmatrix}. \quad (14)$$

From the same pure endmembers on the top of Fig. 6(a), blue points of Fig. 6(c) represent the same 2D slice generated under the BMM model.

As for the **Highly Concentrating Model**, we choose to rely on the same Dirichlet Distribution as in LMM, however with large concentration parameters $\alpha_i \gg 1$ (in practice we used $\alpha_i = 50, \forall i$). This ensures that generated abundance coefficients \mathbf{g}_i are highly concentrated in the middle of the simplex which is formed by the endmembers. This is visible in blue points of Fig. 6(c).

³These signatures together with additional information regarding them can be seen in the supplement material.

In order to generate and include a number \mathcal{Y} of outliers to the nominal synthetic data, we used three spectral signatures $\mathbf{E}' = [\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3]$ without any similarities to the signatures used to generate nominal data. From ℓ known endmembers in \mathbf{E} and our abnormal spectral signatures in \mathbf{E}' , we use again Dirichlet distribution function to generate fractional abundances for anomalies \mathbf{g}'_i using three more parameters $\alpha' = [\alpha'_1, \alpha'_2, \alpha'_3]$. The generation function reads:

$$\mathbf{g}'_i = \frac{\Gamma(\sum_{j=1}^{\ell+3} \gamma_j)}{\prod_{j=1}^{\ell+3} \Gamma(\gamma_j)} \prod_{j=1}^{\ell+3} \mathbf{e}_j^{\gamma_j-1} \quad (15)$$

where $\gamma = [\alpha_1, \dots, \alpha_\ell, \alpha'_1, \alpha'_2, \alpha'_3]$ is the concatenation of parameters α and α' . In practice, we used $\alpha_1, \dots, \alpha_\ell = 1$ and $\alpha'_1, \alpha'_2, \alpha'_3 = 50$ to simulate outliers. These latter are depicted in red in figures 6(b-c-d) and an example of anomaly is visible in the bottom of figure 6(a). As shown on these figures, outliers are lying in a separable distance from the 2-simplex encompassing the nominal data (note that in practice outliers may not necessarily live in a similar area but we prefer to use this simple situation for validation).

In further experiments on synthetic data we use different datasets with various sizes ($N = \{10^3, 10^4, 10^5, 10^6\}$) and endmembers ($\ell = \{3, 5, 7, 10, 12, 15\}$). As for the evaluation of anomaly detection procedures, for all datasets we choose to generate $\mathcal{Y} = 20$ anomalies from three different signatures.

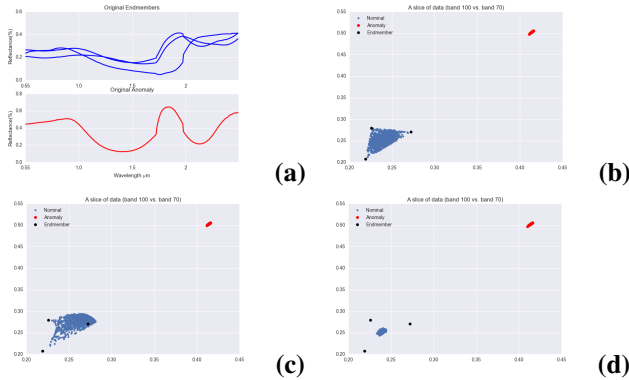


Fig. 6: Illustration of synthetic data. (a): Some endmembers (top) and the anomaly spectrum (bottom) used to generate the datasets. (b): A 2D Slice of two components of a synthetic dataset of 1010 samples generated based on LMM from the 3 nominal endmembers and 10 samples issued from the single anomaly signature. The geometrical location of the anomalies in the 2D plan makes them visually separable. (c): Similar visualization but mixtures are generated on the basis of BMM. (d): Similar visualization but mixtures are generated on the basis of HCM.

3) *Real data:* For the application on **real data** we choose in this study to focus on 4-Vesta asteroid which is the second largest asteroid in the solar system and thus is of prime scientific interest in planetary exploration. 4-Vesta is thought to be a protoplanet, i.e. a building block of terrestrial planets remnant of the early formation of the Solar System [51].

Therefore understanding its properties is of prime importance to understand the initial material that accreted to form the Earth⁴. The number “4” in the name of the asteroid is due to the fact that 4-Vesta was the fourth asteroid to be discovered. It was observed recently by the Dawn mission [52], providing high resolution hyperspectral images of the whole surface. Moreover a very large number of meteorite samples in our collection come from this asteroid [53], providing additional constraints on the minerals mixtures possibly encountered when processing the HS dataset. As the dataset available for 4-Vesta is concerned by all the limitations described in section I, the use of an improved non-linear unmixing algorithm is essential to progress from basic indexes maps (as done up to now, e.g. [54]) to detailed maps, making the link between meteorites and compositions extracted from remote sensing possible.

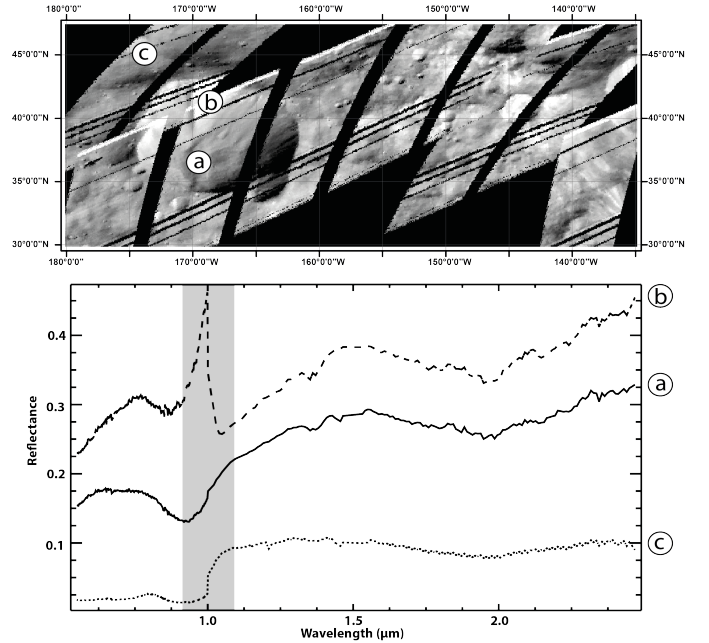


Fig. 7: 4-VESTA mosaic composed of (842×327) pixels, with 383 spectral bands covering the range $0.55\mu\text{m}$ to $2.47\mu\text{m}$. Examples of spectra (located on the mosaic with letters) are reported with (a) a normal spectrum, (b) discontinuity at the two detectors boundary and (c) a spectrum at the boundary of a shadow region. The gray wavelength range is where inconsistency are usually observed.

Because of the large size of the full dataset (as mentioned in section I), here we choose to focus on a small subset located in the northern hemisphere. This region was chosen because Olivine, a mineral indicative of deep magmatic processes, was first detected in this area [55]. This mineral is never found alone in meteorites coming from Vesta-4 [56], [57], implying the use of an algorithm able to decompose spectra as a combination of mixtures of materials.

⁴In astrophysics, accretion is the growth of particles into a massive object by gravitationally attracting more matter, typically gaseous matter and dust, in an accretion disc.

Following the procedure described in [58], we processed Dawn VIR [59] images from the High Altitude Mapping Orbit (HAMO) 1 and 2 and from the Low Altitude Mapping Orbit (LAMO). Raw images were downloaded from PDS in level 1B and were calibrated using ISIS3 pipeline [60] and photometric parameters found in the literature [61]–[63]. Additional in-house routines were used to filter bad pixels and to correct the geographical misalignment between the two detectors of the VIR instrument (following the method described in [54]). Finally each image was converted from radiance to I/F and projected using the spacecraft geometry information. Such process remains standard in planetary image analysis.

After removing channels known to be noisy, the final mosaic has a size of (842×327) pixels, with 383 spectral bands covering the range $0.55\mu\text{m}$ to $2.47\mu\text{m}$. Nevertheless, despite the first-order filtering of bad pixels/spectra, some inconsistency still exist in the dataset and outliers can be observed, as illustrated in Fig. 7. In this figure the typical shape of an expected spectra is depicted together with highly corrupted spectra due to the misalignment between the visible and near-infrared detectors, either showing a sharp peak when pixels fall on two distinct lithological units or an Heaviside-step shape signature when pixels fall at a boundary between sunlit and shadow.

4) *Parameters setting for SAGA and SAGA+*: To guarantee an optimal performance, internal parameters of SAGA and SAGA+ have to be tuned according to dataset characteristics. Apart from the number of atoms/endmembers ℓ and the sparsity level λ , these parameters correspond to kernel bandwidth (σ) and for SAGA+, to the error threshold (τ) of the anomaly detection procedure (see algorithm in Fig. 4). For this reason a number of tests have been taken in order to find the best parametrization of the models. As the result of these tests, we observed empirically that the best parameter values for linear mixtures are $\sigma = 0.15$ and $\tau = 0.2$, for bilinear mixtures : $\sigma = 0.15$ and $\tau = 0.04$ and for highly concentrating mixtures : $\sigma = 0.15$ and $\tau = 0.2$. This setting enabled us to detect all true anomalies (true positives) through each run of the algorithm while keeping the number of wrongly selected anomalies (false positives) minimum. These parameters kept fix in all our experiments.

It should be noted that these tests are meant to find the best parameters for a model regarding the dataset under consideration and thus not to be thought as generic. For the real application presented later we also use $\sigma = 0.15$ and $\tau = 0.2$ since they are used for LMM and HCM. As will be shown in the next section, this setting produces acceptable results. The question of an optimal parameter selection for real datasets within the unsupervised scenario however remains an open question. We now turn to experiments.

B. Results on synthetic data

In this part we compare the results from SAGA and SAGA+ with other unmixing procedures on the synthetic datasets, presented in section III-A2. We first compare SAGA with N-

Finder, VCA, and G-SiVM in order to evaluate their accuracy in endmember estimation in absence of any anomaly.

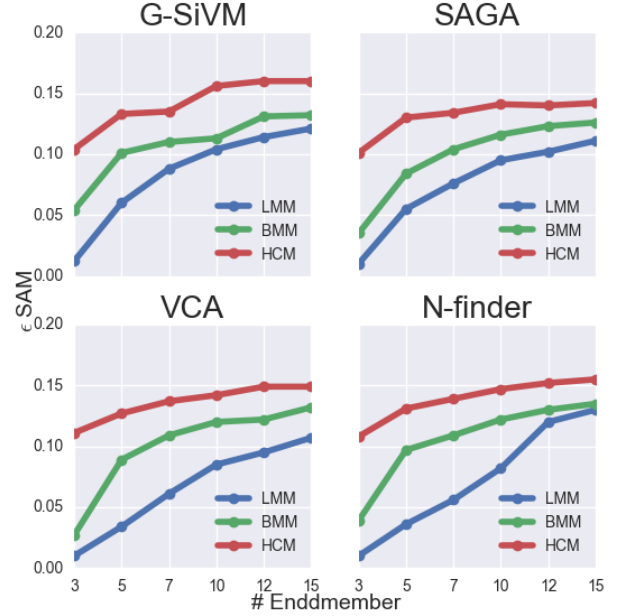


Fig. 8: **Endmember estimation** Averages of SAM values for SAGA, G-SiVM, VCA and N-Finder in function of the number of endmembers. Each plot represents the results for three mixing models.

For this we generated various datasets of size $N = 10^3$, using different numbers of signatures $\ell = 3, 5, 7, 10, 12, 15$ and based on the three aforementioned mixing models. Because our approach embeds some randomness in the initialization step, we depict the average of SAM errors computed on 100 runs. They are visible in Fig. 8 in function of different number of endmembers for each algorithm. From this figure one observes that SAGA, N-Finder and VCA have similar performances unlike G-SiVMs whose accuracy is generally diminished for the case of intimate mixtures. This first experience where no outliers are involved shows the relative good performances of all techniques to extract consistent endmembers.

In a second experience we ran exactly the same experiments, this time however, we included $\mathcal{T} = 20$ anomalies (generated from 3 extra signatures randomly selected from the dataset as explained in previous section) in the dataset. New results are depicted in Fig.9. As expected the overall error increases in all the algorithms except SAGA+ which includes an anomaly detector process. These simple experiments demonstrate the ability of our approach to accurately estimate endmembers while removing outliers. It should be noticed that the 41 signatures used in this setting are not extremely different in their overall profile shape. Nevertheless the results show that the presence of AD even at the situations where the anomalies are not dramatically different (therefore have small distance w.r.t the nominal data) from the nominal signatures can effect

considerably.

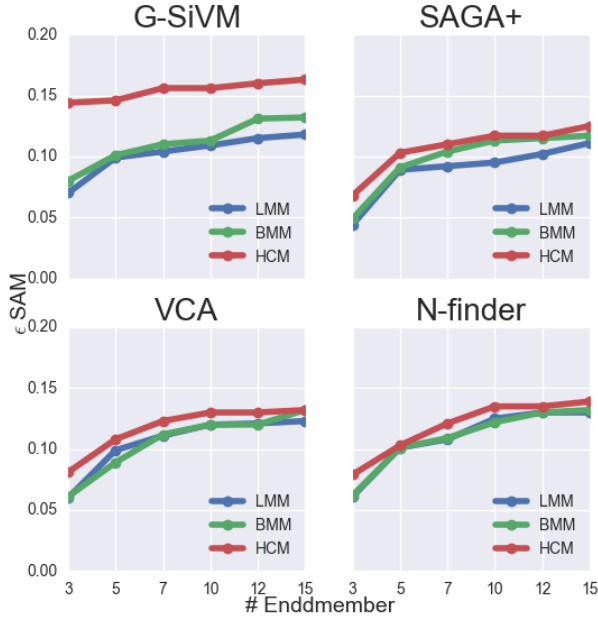


Fig. 9: **Endmember estimation** Averages of SAM values for SAGA+, G-SiVM, VCA and N-Finder in function of the number of endmembers. Data contains anomalies generated based on 3 random signatures. Each plot represents the results for three mixing models.

Let us also outline that in the last dataset, outlier signatures were considerably different from the nominal data, making it an ideal condition for the anomaly detection. In order to evaluate the sensitivity of SAGA+ when anomalies are more dispersed w.r.t data, several concentration levels (issued from Dirichlet distribution as mentioned in previous section) of anomalies have been generated: the higher the concentration level, the lower the dispersion of anomalies. An example of dataset is shown in Fig. 10 together with endmember and anomaly estimation with SAGA+. As one observes in this figure, both anomalies and endmembers are accurately detected.

In another test, the anomalies are assigned on the basis of 50 executions, where in each execution, a new concentration level in the range $\alpha' = 1, \dots, 50$ is used to generate anomalies (see equation (15)). In order to quantitatively evaluate the sensitivity of AD procedure to the level of concentration/dispersion of anomalies. In Table. I we report the average of Kappa values when comparing true number of anomalies to those detected by SAGA+ over synthetic datasets of size $N = 10^3$, generated with various numbers of endmembers $\ell = 3, 5, 7, 10, 12, 15$ and $\Upsilon = 20$ anomalies generated from three different signatures. From this table one observes that when increasing the number of endmembers, the anomaly detection algorithm is still robust since $\kappa > 0.7$ is generally interpreted as a good separation accuracy in two class classification. The way Kappa reacts in function of the concentration of outliers is visually illustrated in Fig. 11.

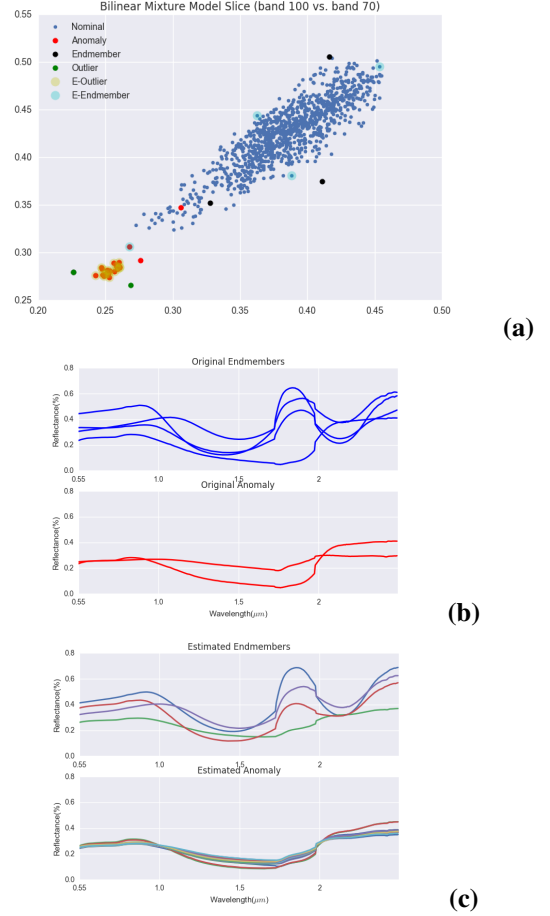


Fig. 10: **Results in presence of anomalies using SAGA+.** (a): Representation of the dataset; (b) original endmembers and anomalies (c) estimated endmembers and anomalies with SAGA+

TABLE I: Kappa statistics for SAGA+ in endmember estimation for the three synthetic datasets

	ℓ	3	5	7	9	11	13	15
κ	LMM	0.73	0.84	0.85	0.84	0.84	0.84	0.79
	BMM	0.94	0.94	0.89	0.88	0.85	0.85	0.78
	IMM	0.96	0.96	0.95	0.95	0.94	0.94	0.91

Let us remind that when α' grows, the separation between outliers and nominal data is more clear. From this figure one observes that when $\alpha' = 15$, we reach $\kappa \simeq 0.8$ which is a very good accordance. In order to provide a more visible understanding of the effect of anomaly dispersion on the accuracy of the AD algorithm, in Fig.12 we generated different datasets using the same endmember signatures and anomaly signatures. Like other tests 1000 samples generated using $\ell = 3$ endmembers and $\Upsilon = 20$ anomalies generated using 2 spectral signatures. For generation of anomalies different levels of concentration have set which are reflected in the parameter $\alpha = [5, 15, 50]$.

The result of the AD algorithm is then evaluated according to the ratio between *True Positive (TP)* and *False Positive (FP)*.

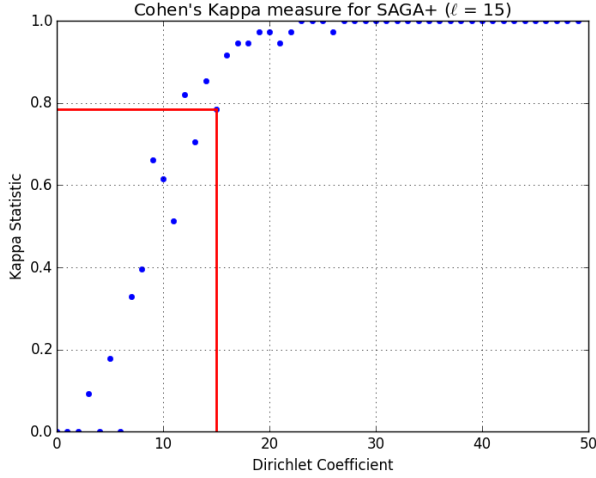


Fig. 11: **Evaluation of SAGA+:** Kappa values as the function of Concentration of the anomalies. The red lines represent the threshold where the Kappa value starts to get over 0.8 which is equal to the concentration level $\alpha = 15$.

It is visible that how large dispersion (low values of α) can obscure the division line between nominal data and outliers, resulting in low accuracy of anomaly detection. In the contrary high value of α isolates the anomalies and thus it would be easier for the AD algorithm to detect them.

C. On the application of Overcomplete Dictionary

In order to illustrate the benefits of overcomplete dictionaries in estimating reliable endmembers, we generated a dataset composed of $N = 10^3$ nominal data on the basis of BMM using 4 original signatures. A number of $T = 20$ anomalies have been added from 2 different signatures. When running SAGA+ we choose $\ell = 10$ atoms to represent endmembers on an overcomplete dictionary and we set a sparsity level $\lambda = 4$ since mixing were generated from 4 signatures. Results are shown in Fig. 13 where we depicted original endmembers and outliers (Fig. 13(a)), a 2D slice of the generated dataset (Fig. 13(b)), the $\ell = 10$ estimated endmembers and associated outliers (Fig. 13(c)) and the resulting 4 “averaged” endmembers (Fig. 13(d)). In this last sub-figure, as we know exactly the location of the 4 pure endmembers, each of the ten estimated ones is associated to the closest real one and weighted averages, on the basis of their abundances, are represented. One can see the good adequacy of our approach when comparing these averaged endmembers with real ones in the top of Fig. 13(a).

It is also interesting to see the effect of overcomplete dictionary on the accuracy of the abundance coefficients. For this we generated datasets in a similar way as the previous step. For the projection of dataset on the estimated endmembers by N-finder and VCA algorithms we used the *Constrained Least Squares* (CLS) [11] method. This time we evaluated the closeness of each row of the estimated abundance matrix with its closest counterpart in the original generated abundance matrix using the mean of *SAM* error calculated as:

$$\epsilon(\mathbf{G}, \hat{\mathbf{G}}) = \min_{\eta} \frac{1}{\ell} \sum_{i=1}^{\ell} \text{SAM}(\mathbf{g}_{\eta^{\ell}(i)}, \hat{\mathbf{g}}_i) \quad (16)$$

where ℓ is the number of endmembers and $\eta^{\ell} : \{1, \dots, \ell\} \rightarrow \{1, \dots, \ell\}$ is a function corresponding to all possible permutations. The result of this comparison is represented in the table II which shows considerable improvement in the accuracy of the abundance maps in the case of SAGA+ comparing to the other methods.

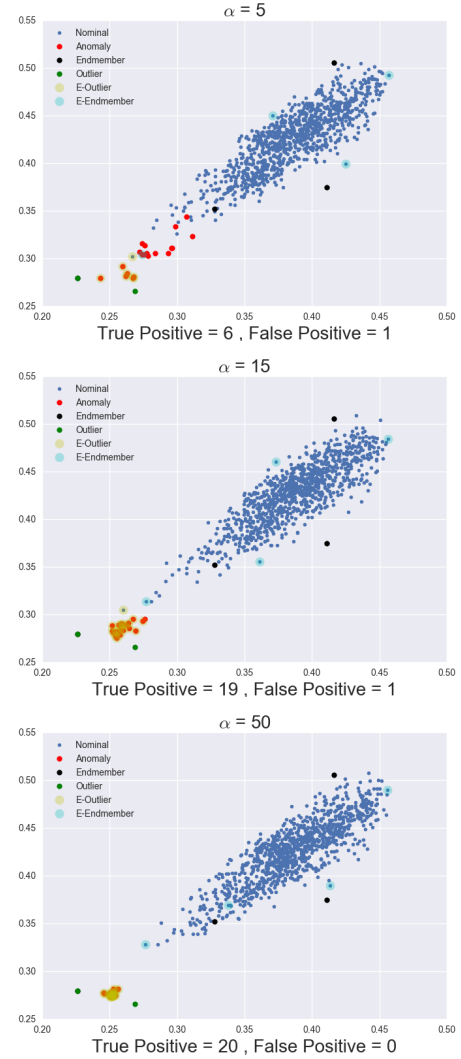


Fig. 12: **Effect of Anomaly concentration:** In each plot a 2D slice of a dataset generated based on the same endmember signatures and anomaly signatures is illustrated. For construction of anomalies different levels of concentration is set : $\alpha = [5, 15, 50]$.

Finally, we report a comparison of execution times on supplementary material.

D. Results on real data: 4-Vesta Asteroid

In this experiment we consider the performance of SAGA+ on 4-Vesta and try to give a high level interpretation of

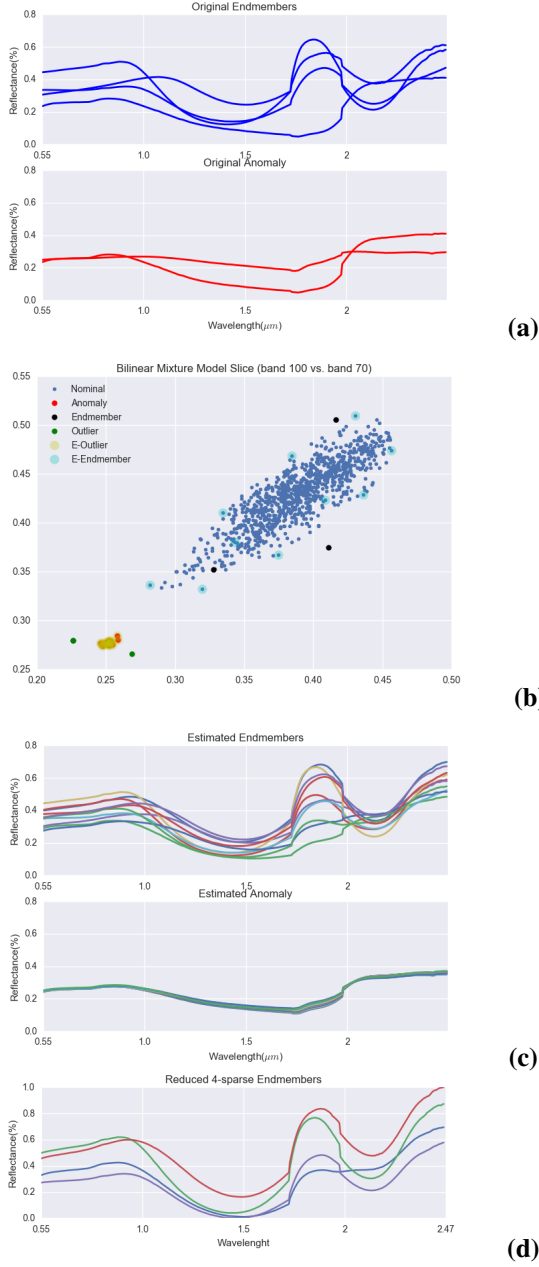


Fig. 13: **Results in presence of anomalies using SAGA+.** (a): original signatures used for generation of dataset; (b): 2D Slice of dataset together with the estimated prototypes (c): signatures of estimated endmembers and anomalies with SAGA+ (d): the 4 weighted averaged endmembers.

TABLE II: Average SAM errors between the original and estimated abundance coefficients. Best results are reported in boldfont

	ℓ	3	5	7	10	12	15
ϵ	VCA	0.237	0.541	0.756	0.912	0.981	1.105
	N-finder	0.155	0.451	0.713	0.918	0.970	1.108
	G-SiVM	0.333	0.592	0.951	1.084	1.121	1.301
	SAGA+	0.096	0.123	0.311	0.549	0.711	0.892

the results which can be of interest from the planetary observation point of view. In practice, two difficulties arise when manipulating 4-Vesta data : 1) the real number of endmembers is unknown and 2) the assumption that some pure pixels are present in the original image does not hold. In order to cope with these difficulties, as mentioned earlier, we exploit the intrinsic characteristics of *overcomplete* dictionary learning approach which relaxes both mentioned difficulties by selecting out a large number of endmembers (referred to as the dictionary atoms) from the image.

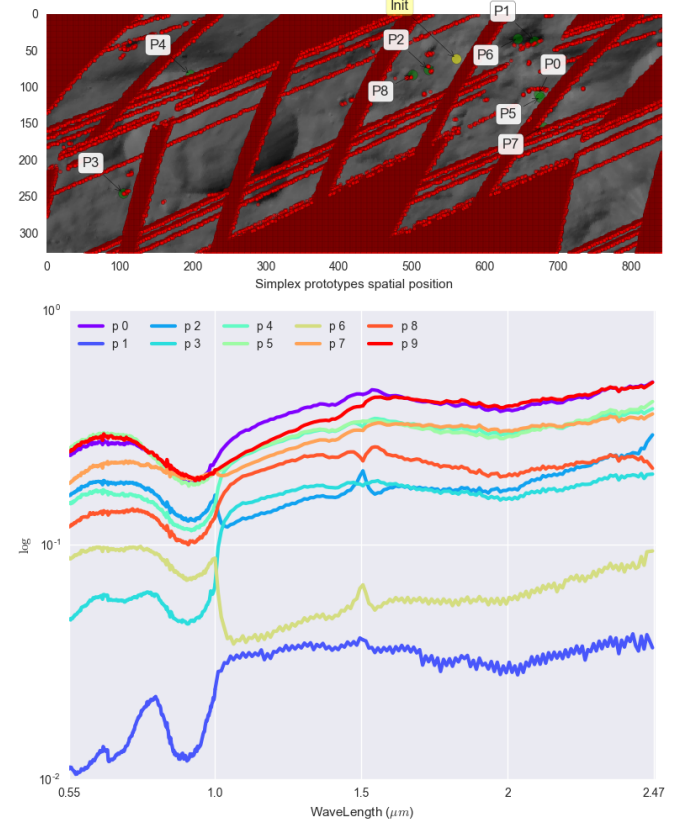


Fig. 14: **Results on real 4-Vesta asteroid.** **Top:** Vesta-4 mosaic with the spatial position of the extracted endmembers. The red points are detected anomalies and the 10 green points with labels are selected endmember prototypes. **Bottom:** The respective estimated endmember signatures.

From the geological point of view, the large number of selected endmembers/atoms might be interpreted as different compositions of materials in a scene. In practice for the case of 4-Vesta we choose the dictionary of size $\ell = 10$. Because of the intrinsic characteristic of overcomplete dictionary for re-grouping the similar endmembers, this parameter is not crucial but has to be set such that the number of atoms of the dictionary is larger than the number of expected “pure” materials.

The estimated endmembers are represented in the bottom of Fig. 14 (denoted P0 to P9) while the top of this figure depicts the spatial positions of the respective estimated endmembers (in green) together with the outliers (in red). A specimen of

outliers spectra signature (with a sharp peak) is represented in the bottom of Fig. 7.

By comparing these two plots (Fig. 14 and Fig. 7), one can observe that the estimated endmembers signatures are generally in accordance with the expected nominal signature profiles (laboratory spectra related to the lithologies on the northern hemisphere). Although one observes some unwanted signatures (mainly contributing to the shadowing effect), it is of prime interest to note that the absence of signatures corrupted by noise is significant⁵. Indeed, endmembers related to unwanted spectra can be discarded: P1 and P3 have a low reflectance and correlate perfectly with the most dark (shadow) areas, while P2, P6 and P7 show a shift between the two detectors, with a very small amount of noise. The five other endmembers are of interest regarding mineralogical detections. P4 and P8 show characteristic signatures of pyroxenes and represent the surface background. Some outcrops, related to P0, exhibit similar signatures with stronger absorptions, in relation with impact craters (i.e. fresh rocks) and very good illumination conditions. The last two endmembers P5 and P9, detected in small spots in and around craters, show a broader absorption feature in the $1\mu m$ range.

One can see the whole set of endmembers and their respective abundance maps in Fig. 17. From these plots, it is worth noting that the spatial consistency of associated maps matches well with real physical structures in 4-Vesta, which is a very good property.

As a more physical comparison, one can also compare the spectra in Fig. 14 with the laboratory measured spectra of Olivine and Orthopyroxene which is illustrated in Fig. 15. As can be observed in this figure, the broader $1\mu m$ absorption seen in P9 and P5 can be correlated to an increased olivine content in the rocks, in accordance with observations made by [55] or [64]. From these observations, SAGA+ is thus able to extract from the hyperspectral image spectra that are a combination of mixtures of materials, representative of distinct geological units. Additionally, using the extracted endmembers, the outcrops enriched in Olivine (i.e. spots where pixels in which P9 and P5 are needed in the unmixing) can be mapped (see Fig. 16) and found localization are similar to those obtain with other techniques [64]. Let us note that that the shift in absorption center at $2\mu m$ is due to the chemical composition of orthopyroxene (olivine does not have absorptions in this wavelength range) and reflects the fact that the orthopyroxene used as a reference does not have the same chemical composition as the one observed on 4-Vesta's surface. Lastly, in order to represent the benefits of the anomaly detection procedure, we have applied the same method without the outlier removal process. Extracted endmembers and associated abundances are available in the supplement material. By comparing them with Fig. 17, we clearly observe corrupted endmembers yielding less consistent abundance maps. This experience reveals the importance of the anomaly detection process introduced here.

⁵Interested reader may find a plot of detected anomaly signatures from this mosaic in the supplement material

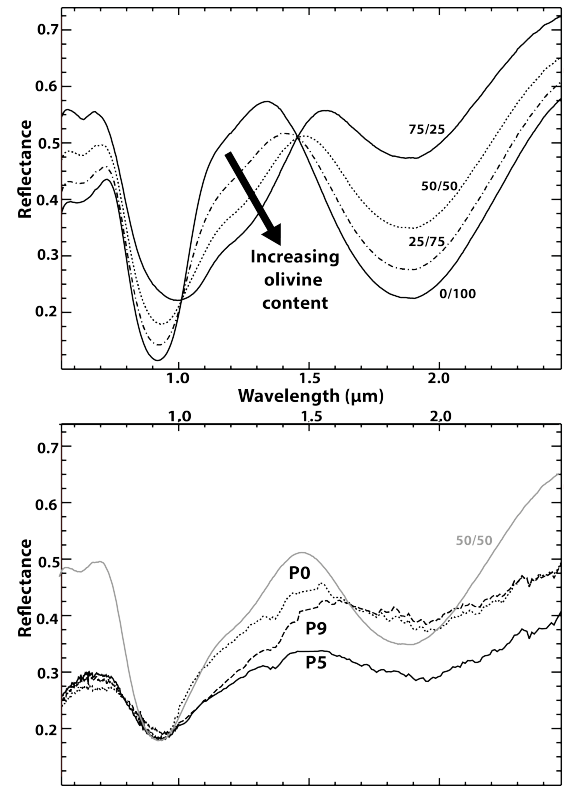


Fig. 15: **Top:** laboratory spectra of mixtures involving olivine and orthopyroxene (spectra from the RELAB library with numbers referring respectively to the olivine and orthopyroxene contents (% olivine / % orthopyroxene). **Bottom:** comparison with the three extracted endmembers P0, P5 and P9. The shift in absorption center at $2\mu m$ is due to an increase in Ca and/or Fe in the pyroxene, compared to the pure orthopyroxene composition.

IV. CONCLUSION

In this paper, a joint procedure for anomaly detection and spectral unmixing has been presented for hyperspectral data. Endmembers and anomalies are extracted from the vertexes of the simplex that embeds data, this simplex being computed in a feature space to deal with non-linearities. Abundances are computed on the basis of sparse projections of spectra onto extracted endmembers. To deal both with the internal heterogeneity inside classes and with the fact that pure elements are not necessarily present in the data, overcomplete dictionaries have been used. We evaluated and compared our approach with linear and non linear unmixing methods on synthetic datasets generated from real spectra where linear and nonlinear mixing procedures have been applied. Anomalies have also been introduced in the data. These experiments have revealed the efficiency of our technique to deal with anomalies and endmember detection at the same time. Computational aspects have also been discussed and have revealed the ability of our algorithm to easily scale up to large datasets. Lastly, experiments have been performed on real data issued from the 4-Vesta Asteroid. From a geological point of view, we have shown that anomalies, endmembers and mixing abundances

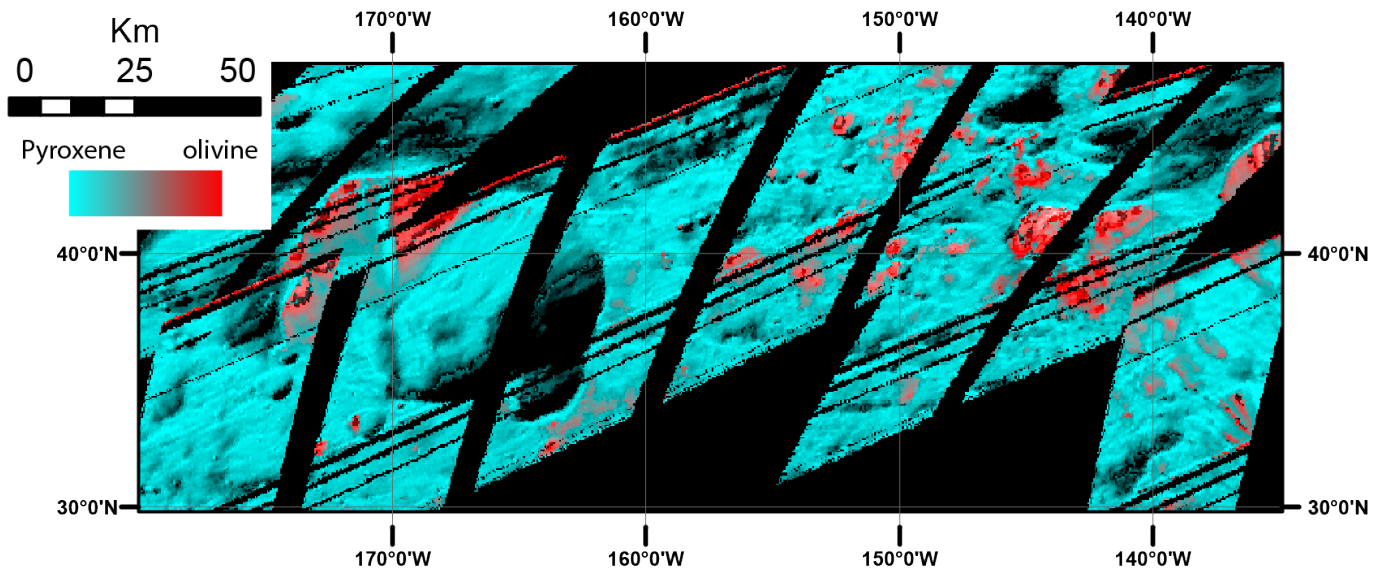


Fig. 16: RGB composition with $R=P9+P5$, $G=P4+P8+P0$, and $B=P4+P8+P0$. Coloring shows the localization of outcrops enriched in olivine (appearing in red)

are consistent, demonstrating the efficiency of our approach for real applications.

ACKNOWLEDGMENT

The Authors would like to thank Conseil Régional de Bretagne and Université de Bretagne sud for their support of the project ARED -RB OPTIMAP. We also thank ANR ASTERIX project for supporting this work. We thank the authors in [11] for the available package and codes provided along the survey. Also we would like to thank the authors in [7], [22], [48] for providing us the codes of their algorithms, which made the comparisons possible within this work.

REFERENCES

- [1] D. Stein, S. Beaven, L. Hoff, E. Winter, A. Schaum, and A. Stocker, "Anomaly detection from hyperspectral imagery," *Signal Processing Magazine, IEEE*, vol. 19, no. 1, pp. 58–69, Jan 2002.
- [2] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *Geoscience and Remote Sensing Magazine, IEEE*, vol. 1, no. 2, pp. 6–36, 2013.
- [3] M. T. Eismann, *Hyperspectral remote sensing*. SPIE Press Bellingham, 2012.
- [4] X. Ceamanos, S. Douté, B. Luo, F. Schmidt, G. Jouannic, and J. Chanussot, "Intercomparison and Validation of Techniques for Spectral Unmixing of Hyperspectral Images: A Planetary Case Study," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4341–4358, 2011.
- [5] J. B. Adams, "Interpretation of visible and near-infrared diffuse reflectance spectra of pyroxenes and other rock forming minerals," in *Infrared and Raman spectroscopy of lunar and terrestrial materials*, C. Karr, Ed. New York Academic Press, 1975, pp. 91–116.
- [6] J.-P. Bibring, M. Combes, Y. Langevin, A. Soufflot, C. Cara, P. Drossart, T. Encrenaz, S. Erard, O. Forni, B. Gondet, L. Ksanfomalft, E. Lelouch, P. Masson, V. Moroz, F. Rocard, J. Rosenqvist, and C. Sotin, "Results from the ISM experiment," *Nature*, vol. 341, pp. 591–593, 1989.
- [7] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model," *Signal Processing, IEEE Transactions on*, vol. 61, no. 2, pp. 480–492, Jan 2013.
- [8] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. Bermudez, S. McLaughlin, and A. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *Signal Processing Magazine, IEEE*, vol. 31, no. 1, pp. 82–94, Jan 2014.
- [9] H. Clenet, P. C. Pinet, Y. Daydou, F. Heuripeau, C. Rosemberg, D. Baratoux, and S. Chevrel, "A new systematic approach using the Modified Gaussian Model: Insight for the characterization of chemical composition of olivines, pyroxenes and olivinepyroxene mixtures," *Icarus*, vol. 213, pp. 404–422, 2011.
- [10] P. Boccacci, *Introduction to Inverse Problems in Imaging*. CRC Press.
- [11] J. Bioucas-Dias and A. Plaza, "An overview on hyperspectral unmixing: Geometrical, statistical, and sparse regression based approaches," in *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, July 2011, pp. 1135–1138.
- [12] J. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 5, no. 2, pp. 354–379, 2012.
- [13] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. Bermudez, S. McLaughlin, and A. O. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *Signal Processing Magazine, IEEE*, vol. 31, no. 1, pp. 82–94, 2014.
- [14] B. Hapke, *Theory of reflectance and emittance spectroscopy*. Cambridge University Press, 2012.
- [15] J. M. P. Nascimento and J. M. Bioucas-Dias, "Nonlinear mixture model for hyperspectral unmixing," pp. 74 770I–74 770I–8.
- [16] W. Fan, B. Hu, J. Miller, and M. Li, "Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data," *International Journal of Remote Sensing*, vol. 30, no. 11, pp. 2951–2962, 2009.
- [17] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 11, pp. 4153–4162, 2011.
- [18] I. Meganem, P. Deliot, X. Briottet, Y. Deville, and S. Hosseini, "Linear-quadratic mixing model for reflectances in urban environments," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 1, pp. 544–558, 2014.
- [19] J. Plaza, A. J. Plaza, P. Martinez, and R. M. Perez, "Nonlinear mixture models for analyzing laboratory simulated-forest hyperspectral data," in *Remote Sensing*. International Society for Optics and Photonics, 2004, pp. 480–487.
- [20] C.-I. Chang and A. Plaza, "A fast iterative algorithm for implementation of pixel purity index," *Geoscience and Remote Sensing Letters, IEEE*, vol. 3, no. 1, pp. 63–67, Jan 2006.

- [21] M. E. Winter, "N-findr: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," pp. 266–275.
- [22] J. Nascimento and J. Bioucas Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 4, pp. 898–910, April 2005.
- [23] M. Craig, "Minimum-volume transforms for remotely sensed data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 32, no. 3, pp. 542–552, May 1994.
- [24] M. Hollósi, G. D. Fasman *et al.*, "Convex constraint analysis: a natural deconvolution of circular dichroism curves of proteins," *Protein Engineering*, vol. 4, no. 6, pp. 669–679, 1991.
- [25] S. Jia and Y. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 47, no. 1, pp. 161–173, 2009.
- [26] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, no. 3, pp. 765–777, 2007.
- [27] P. Sajda, S. Du, and L. Parra, "Recovery of constituent spectra using non-negative matrix factorization," in *Optical Science and Technology, SPIE's 48th Annual Meeting*. International Society for Optics and Photonics, 2003, pp. 321–331.
- [28] N. Yokoya, J. Chanussot, and A. Iwasaki, "Generalized bilinear model based nonlinear unmixing using semi-nonnegative matrix factorization," in *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*. IEEE, 2012, pp. 1365–1368.
- [29] R. Heylen and P. Scheunders, "Calculation of geodesic distances in nonlinear mixing models: Application to the generalized bilinear model," *Geoscience and Remote Sensing Letters, IEEE*, vol. 9, no. 4, pp. 644–648, 2012.
- [30] N. Courty, X. Gong, J. Vandel, and T. Burger, "Saga: sparse and geometry-aware non-negative matrix factorization through non-linear local embedding," *Machine Learning*, vol. 97, no. 1-2, pp. 205–226, 2014.
- [31] J. Broadwater, A. Banerjee, and P. Burlina, "Kernel methods for unmixing hyperspectral imagery," *Kernel Methods for Remote Sensing Data Analysis*, pp. 249–270, 2009.
- [32] J. Broadwater and A. Banerjee, "A comparison of kernel functions for intimate mixture models," in *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. WHISPERS '09. First Workshop on*, Aug 2009, pp. 1–4.
- [33] J. Broadwater, R. Chellappa, A. Banerjee, and P. Burlina, "Kernel fully constrained least squares abundance estimates," in *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, July 2007, pp. 4041–4044.
- [34] J. Bieniarz, E. Aguilara, X. Zhu, R. Muller, and P. Reinartz, "Joint sparsity model for multilook hyperspectral image unmixing," *Geoscience and Remote Sensing Letters, IEEE*, vol. 12, no. 4, pp. 696–700, April 2015.
- [35] J. Bieniarz, R. Muller, X. Zhu, and P. Reinartz, "On the use of over-complete dictionaries for spectral unmixing," in *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2012 4th Workshop on*, June 2012, pp. 1–4.
- [36] J. Li and J. M. Bioucas-dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," 2008.
- [37] T.-H. Chan, W.-K. Ma, A. Ambikapathi, and C.-Y. Chi, "A simplex volume maximization framework for hyperspectral endmember extraction," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 11, pp. 4177–4193, Nov 2011.
- [38] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *Aerospace and Electronic Systems Magazine, IEEE*, vol. 25, no. 7, pp. 5–28, July 2010.
- [39] T. Veracini, S. Matteoli, M. Diani, and G. Corsini, "Fully unsupervised learning of gaussian mixtures for anomaly detection in hyperspectral imagery," in *Intelligent Systems Design and Applications, 2009. ISDA'09. Ninth International Conference on*. IEEE, 2009, pp. 596–601.
- [40] S. Schweizer and J. Moura, "Hyperspectral imagery: clutter adaptation in anomaly detection," *Information Theory, IEEE Transactions on*, vol. 46, no. 5, pp. 1855–1871, Aug 2000.
- [41] M. Z. Baghbi, K. Jamshidi, A. R. N. Nilchi, and S. Homayouni, "Improvement of anomaly detection algorithms in hyperspectral images using discrete wavelet transform," *arXiv preprint arXiv:1201.2025*, 2012.
- [42] L. Chapel and C. Friguet, "Anomaly detection with score functions based on the reconstruction error of the kernel pca," in *Machine Learning and Knowledge Discovery in Databases*, ser. Lecture Notes in Computer Science, T. Calders, F. Esposito, E. Hillermeier, and R. Meo, Eds. Springer Berlin Heidelberg, 2014, vol. 8724, pp. 227–241.
- [43] H. Hoffmann, "Kernel-PCA for novelty detection," *Pattern Recognition*, vol. 40, no. 3, pp. 863 – 874, 2007.
- [44] Y. Altmann, S. McLaughlin, and A. Hero, "Robust linear spectral unmixing using anomaly detection," *IEEE Transactions on Computational Imaging*, vol. 1, no. 2, pp. 74–85, 2015.
- [45] C. Thureau, K. Kersting, M. Wahabzada, and C. Bauckhage, "Descriptive matrix factorization for sustainability adopting the principle of opposites," *Data Mining and Knowledge Discovery*, vol. 24, no. 2, pp. 325–354, 2012.
- [46] G. Camps-Valls and L. Bruzzone, *Kernel methods for remote sensing data analysis*. John Wiley & Sons, 2009.
- [47] A. T. Kyrillidis, S. Becker, and V. Cevher, "Sparse projections onto the simplex," *CoRR*, vol. abs/1206.1529, 2012. [Online]. Available: <http://arxiv.org/abs/1206.1529>
- [48] R. Heylen, D. Burazerovic, and P. Scheunders, "Non-linear spectral unmixing by geodesic simplex volume maximization," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 3, pp. 534–542, June 2011.
- [49] J. Cohen *et al.*, "A coefficient of agreement for nominal scales," *Educational and psychological measurement*, vol. 20, no. 1, pp. 37–46, 1960.
- [50] F. Schmidt, A. Schmidt, E. Tréguier, M. Guheneuf, S. Moussaoui, and N. Dobigeon, "Implementation strategies for hyperspectral unmixing using bayesian source separation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4003–4013, 2010.
- [51] C. T. Russell, C. A. Raymond, A. Coradini, H. Y. McSween, M. T. Zuber, A. Nathues, M. C. De Sanctis, R. Jaumann, a. S. Konopliv, F. Preusker, S. W. Asmar, R. S. Park, R. Gaskell, H. U. Keller, S. Mottola, T. Roatsch, J. E. C. Scully, D. E. Smith, P. Tricarico, M. J. Toplis, U. R. Christensen, W. C. Feldman, D. J. Lawrence, T. J. McCoy, T. H. Prettyman, R. C. Reedy, M. E. Sykes, and T. N. Titus, "Dawn at Vesta: testing the protoplanetary paradigm." *Science*, no. 6082, pp. 684–686, May.
- [52] C. T. Russell and C. A. Raymond, "The Dawn Mission to Vesta and Ceres," *Space Science Reviews*, no. 1-4, pp. 3–23, Oct.
- [53] H. Y. McSween, R. P. Binzel, M. C. De Sanctis, E. Ammannito, T. H. Prettyman, A. W. Beck, V. Reddy, L. Le Corre, M. J. Gaffey, T. B. McCord, C. A. Raymond, and C. T. Russell, "Dawn; the Vesta-HED connection; and the geologic context for eucrites, diogenites, and howardites," *Meteoritics & Planetary Science*, no. 11, pp. 2090–2104, Nov.
- [54] E. Ammannito, M. C. De Sanctis, F. Capaccioni, M. T. Capria, F. Carraro, J.-P. Combe, S. Fonte, A. Frigeri, S. P. Joy, A. Longobardo, G. Magni, S. Marchi, T. B. McCord, L. A. McFadden, H. Y. McSween, E. Palomba, C. M. Pieters, C. A. Polanskey, C. A. Raymond, J. M. Sunshine, F. Tosi, F. Zambon, and C. T. Russell, "Vestan lithologies mapped by the visual and infrared spectrometer on Dawn," *Meteoritics & Planetary Science*, no. 11, pp. 2185–2198, Sep.
- [55] E. Ammannito, M. C. De Sanctis, E. Palomba, A. Longobardo, D. W. Mittlefehldt, H. Y. McSween, S. Marchi, M. T. Capria, F. Capaccioni, A. Frigeri, C. M. Pieters, O. Ruesch, F. Tosi, F. Zambon, F. Carraro, S. Fonte, H. Hiesinger, G. Magni, L. A. McFadden, C. A. Raymond, C. T. Russell, and J. M. Sunshine, "Olivine in an unexpected location on Vesta's surface," *Nature*, pp. 122–125.
- [56] A. W. Beck and H. Y. McSween, "Diogenites as polymict breccias composed of orthopyroxenite and harzburgite," *Meteoritics & Planetary Science*, no. 5, pp. 850–872, Aug.
- [57] J.-P. Combe, T. B. McCord, L. A. McFadden, S. Ieva, F. Tosi, A. Longobardo, A. Frigeri, M. C. De Sanctis, E. Ammannito, O. Ruesch *et al.*, "Composition of the northern regions of vesta analyzed by the dawn mission," *Icarus*, vol. 259, pp. 53–71, 2015.
- [58] H. Clenet, M. Jutzi, J.-A. Barrat, E. I. Asphaug, W. Benz, and P. Gillet, "A deep crust-mantle boundary in the asteroid 4 Vesta," *Nature*, no. 7509, pp. 303–306, Jul.
- [59] M. C. De Sanctis, A. Coradini, E. Ammannito, G. Filacchione, M. T. Capria, S. Fonte, G. Magni, A. Barbis, A. Bini, M. Dami, I. Ficai-Verltoni, and G. Preti, "The VIR Spectrometer," *Space Science Reviews*, no. 1-4, pp. 329–369, Oct.
- [60] J. Anderson, K. J. Becker, T. N. Titus, M. C. De Sanctis, A. Nathues, F. Tosi, S. Schroeder, L. Le Corre, M. S. Kelley, C. A. Raymond, and C. T. Russell, "Isis cartographic tools for the Dawn Framing Camera and Visual and Infrared Spectrometer," in *AGU Fall Meeting*. American Geophysical Union, 2011, pp. #U31A-0009.
- [61] V. Reddy, A. Nathues, L. Le Corre, H. Sierks, J.-Y. Li, R. Gaskell, T. J. McCoy, A. W. Beck, S. E. Schröder, C. M. Pieters, K. J. Becker, B. J.

- Buratti, B. W. Denevi, D. T. Blewett, U. Christensen, M. J. Gaffey, P. Gutierrez-Marques, M. Hicks, H. U. Keller, T. Maue, S. Mottola, L. A. McFadden, H. Y. McSween, D. W. Mittlefehldt, D. P. O'Brien, C. Raymond, and C. T. Russell, "Color and albedo heterogeneity of Vesta from Dawn," *Science*, vol. 336, no. 6082, pp. 700–704, 2012.
- [62] M. C. De Sanctis, E. Ammannito, M. T. Capria, F. Capaccioni, J.-P. Combe, A. Frigeri, A. Longobardo, G. Magni, S. Marchi, T. B. McCord, E. Palomba, F. Tosi, F. Zambon, F. Carraro, S. Fonte, Y. J. Li, L. A. McFadden, D. W. Mittlefehldt, C. M. Pieters, R. Jaumann, K. Stephan, C. A. Raymond, and C. T. Russell, "Vesta's mineralogical composition as revealed by the visible and infrared spectrometer on Dawn," *Meteoritics & Planetary Science*, no. 11, pp. 2166–2184, Jul.
- [63] J.-Y. Li, B. J. Buratti, F. Capaccioni, M. T. Capria, L. Le Corre, B. W. Denevi, M. C. De Sanctis, H. Hoffmann, M. D. Hicks, L. Jorda, H. U. Keller, N. Mastrodemos, S. Mottola, A. Nathues, C. M. Pieters, V. Reddy, C. A. Raymond, T. Roatsch, C. T. Russell, S. E. Schröder, M. V. Sykes, and T. N. Titus, "Photometric properties of Vesta," in *Asteroids, Comets, Meteors*, 2012, p. 6387.
- [64] O. Ruesch, H. Hiesinger, M. C. De Sanctis, E. Ammannito, E. Palomba, A. Longobardo, F. Zambon, F. Tosi, M. T. Capria, F. Capaccioni, A. Frigeri, S. Fonte, G. Magni, C. a. Raymond, and C. T. Russell, "Detections and geologic context of local enrichments in olivine on Vesta with VIR/Dawn data," *Journal of Geophysical Research: Planets*, vol. 119, pp. 1–31, aug 2014.

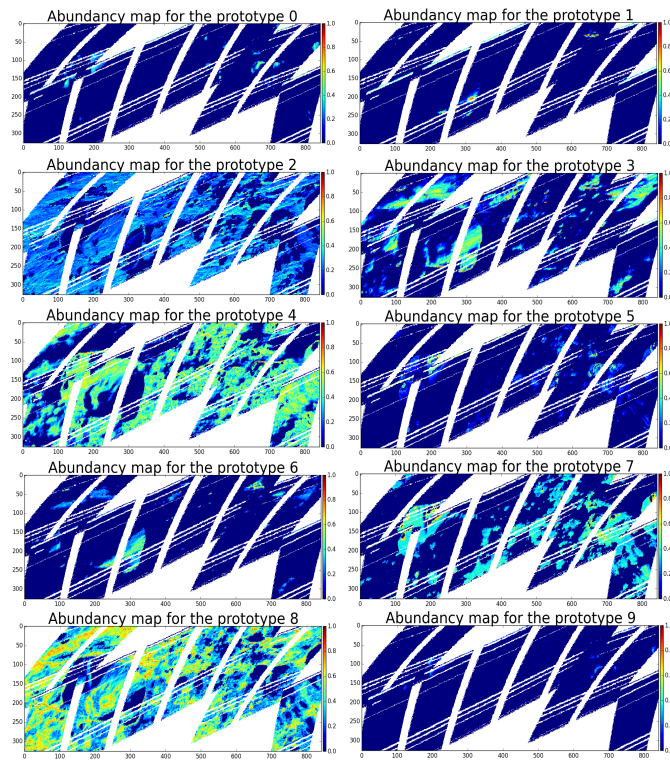


Fig. 17: Abundance maps and respective spectral signatures extracted by SAGA+ (with anomaly detection procedure). Using an anomaly detection procedure, one observes that signatures with magnitudes higher than 1 and those with sharp peaks are not selected as endmembers.



Sina Nakhostin received his M.S. in Robotics and Intelligent Systems from the Örebro University, Sweden in 2013. Since then He has been a Ph.D candidate and member of Obelix team at IRISA laboratory in Université de Bretagne-Sud in Vannes, France. His research interests are Machine Learning and Image processing with a particular interest on Hyperspectral remote sensing analysis.



Harold Clenet received the Ph.D. degree in Planetary Sciences from the Université Paul Sabatier, Toulouse, France, in 2009. His Ph.D. work was on the processing of airborne and orbital hyperspectral images, developing a non-linear algorithm to characterize the composition of magmatic rocks, with applications to the study of the Earth and Mars. For three years, he was a Postdoctoral researcher with the Laboratoire de Géologie de Lyon, Ecole Normale Supérieure de Lyon-Université Claude Bernard Lyon 1, Lyon, France, where he worked with high resolution hyperspectral images from the Compact Reconnaissance Imaging Spectrometer for Mars/Mars Reconnaissance Orbiter (CRISM/MRO) experiment to characterize the composition of the early Martian crust. He also worked on lunar images using the Moon Mineralogy Mapper (M³) instrument on board the Chandrayaan 1 mission. He is currently a Researcher in planetary sciences with the Laboratoire de Sciences de la Terre et des Plantes, Ecole Polytechnique Fédérale de Lausanne, Suisse. His research focus on hyperspectral remote sensing of various objects in the Solar System, with emphasis on asteroid Vesta using data from the Dawn mission (VIR imaging spectrometer). He is also Science Leader for phase 0 of SOLVE (Satellite Observing of Lakes and Vegetation Environments), an Earth observation mission which use a constellation of small satellites to monitor water and vegetation at hourly timescale.



Thomas Corpetti is a Senior Researcher with CNRS (French national center for scientific research). His main research interests concern remote sensing image processing for environmental applications. He has been mainly involved in projects related to the integration of physical models in computer vision tools, in particular using data assimilation, with applications related to meteorology (wind estimation, turbulence analysis, super-resolution) & crop monitoring (plant growth estimation), the long term objective being to analyze links between land cover and atmosphere. At the moment, under the same goal, his research interests mainly concern unmixing and urban vegetation monitoring with various high resolution sensors (LiDAR, PLEIADES, Sentinel-2). He works with LETG (littoral, environment, remote sensing and geomatics) Lab and from 2009 to 2012, he was in Beijing, China, working with LIAMA (sino-french lab on computer sciences and applied mathematics).



Nicolas Courty is associate professor within University Bretagne-Sud since October 2004. He obtained his habilitation degree (HDR) in 2013. His main research objectives are data analysis/synthesis schemes, machine learning and visualization problems, with applications in computer vision, remote sensing and computer graphics. In 2012, he was an Invited professor with a Senior Chinese Academy of Science (CAS) fellowship at the Institute of Automation (CASIA), Beijing, China. In 2014, he spent two months at LASIG (EPFL) as an invited professor from EPFL. Nicolas serves as a regular reviewer for international journals like IEEE JSTARS, TGRS, Data mining and Knowledge Discovery or Machine Learning. Visit <http://people.irisa.fr/Nicolas.Courty/> for more information.